



LAWRENCE  
LIVERMORE  
NATIONAL  
LABORATORY

# Leveraging Genomics Software to Improve Proteomics Results

I.K. Fodor, D.O. Nelson

September 8, 2005

## Disclaimer

---

This document was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor the University of California nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or the University of California. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or the University of California, and shall not be used for advertising or product endorsement purposes.

This work was performed under the auspices of the U.S. Department of Energy by University of California, Lawrence Livermore National Laboratory under Contract W-7405-Eng-48.

# Leveraging genomics software to improve proteomics results

IK Fodor and DO Nelson

September 6, 2005

## Abstract

Rigorous data analysis techniques are essential in quantifying the differential expression of proteins in biological samples of interest. Statistical methods from the microarray literature were applied to the analysis of two-dimensional difference gel electrophoresis (2-D DIGE) proteomics experiments, in the context of technical variability studies involving human plasma. Protein expression measurements were corrected to account for observed intensity-dependent biases within gels, and normalized to mitigate observed gel to gel variations. The methods improved upon the results achieved using the best currently available 2-D DIGE proteomics software. The spot-wise protein variance was reduced by 10% and the number of apparently differentially expressed proteins was reduced by over 50%.

## 1 Motivation

Understanding systems biology is a major scientific challenge. While the genomes of several species have been fully sequenced, the biochemical functions of many genes have not yet been determined. Genes encode proteins, which in turn perform all biochemical functions necessary for life. After genomics (the study of genes in an organism), proteomics (the study of proteins and their functions) is a next, and more difficult, step in understanding cellular biology and the mechanism of diseases.

Two-dimensional difference gel electrophoresis (2-D DIGE) is a leading proteomics technology that investigates simultaneously the proteomes of different biological samples. The proteins that are differentially expressed in the different samples hold the biological clues of interest. The proteomics team within the Biodefense group

in the Biosciences directorate at LLNL led by Dr. Sandra L. McCutchen-Maloney is at the forefront of 2-D DIGE research. They have conducted the largest and most complex experiments in the world on this platform to date (according to our literature survey), and they have a large number of future investigations planned. Their current experiments involve mouse and ex vivo human samples exposed to pathogens such as *Bacillus anthracis* (the causative agent of anthrax) and *Yersinia pestis* (the causative agent of plague). They use DeCyder, the best commercially available product (GE Healthcare), to analyze their experiments. However, they are concerned that the limited statistical techniques in DeCyder are not adequate for analyzing rigorously their experiments.

The analytical tools in DeCyder permit rudimentary statistical analysis, but they have several shortcomings. For example, the system was designed for small studies, not the large throughput experiments characteristic to experiments carried out in the Biosciences directorate. In addition, the data normalization methods and the quantification of the differential expression of the proteins could be improved by incorporating recent advances in statistical techniques. Genomics datasets from microarray chip experiments have encountered similar normalization and expression quantification problems in the past. Since genomics is a more mature field than proteomics, open source software for analyzing microarrays has already incorporated the improved statistical methodology.

This project introduced state-of-the-art software to the problem of analyzing 2-D DIGE in CAR and the Biosciences directorate. The major steps of the solution approach involved:

- Obtaining 2-D gels from biological experiments
- Analyzing the gels with the best commercially available proteomics software

- Analyzing the gels with the open source microarray analysis software
- Comparing the results of the two previous steps and quantifying the differences

## 2 Description of 2-D DIGE

Two-dimensional polyacrylamide gel electrophoresis (2-D PAGE) is a technology by which thousands of proteins in a biological sample are separated according to their isoelectric points and molecular weights [10, 6, 9]. In theory, each protein is uniquely determined by its response along the two dimensions of separation. Differences in the proteomes of multiple samples can be studied by comparing the expression profiles of the proteins on the gels. In traditional 2-D PAGE, each gel contains one sample which is compared to samples on different gels, introducing high experimental variability.

Two-dimensional difference gel electrophoresis (2-D DIGE) was proposed in [14] as a method to overcome gel to gel variability inherent in 2-D PAGE. More recently, 2-D DIGE has been commercialized through the Ettan DIGE System of Amersham Biosciences (now part of GE Healthcare) thanks to the development of the three size and charge-matched, spectrally resolvable CyDye fluors Cy2, Cy3 and Cy5. Gels using the DIGE method contain three samples labeled with the three distinct fluorescent dyes Cy2, Cy3, and Cy5. Typically, two dyes are used to label two different biological samples of interest. The third dye can be used to label the “internal standard” which is a pooled mixture of all the samples used in the experiment, and is identical on all gels. The power of the internal standard is in its potential to adjust for the variability between gels and thus make the data across the experiment more comparable. The DeCyder Differential Analysis Software is part of the Ettan DIGE System, and is used for analyzing the data and quantifying the differential expression of the proteins [13, 1, 2].

Although there are fundamental differences in 2-D DIGE and gene expression microarray technologies, many of the difficulties encountered in the analysis of 2-D DIGE data are similar to problems that arise in the analysis of microarray experiments: proper normalization of the data within and between the gels (arrays), multiple hy-

Table 1: The datasets, as indicated under the Name heading. The Gel column indicates whether the gels were poured in the lab or bought from a vendor. The Filt column specifies whether an exclusion filter was applied to the initial spots detected on the gels. The Landmrk column indicates whether manual landmarking was performed when matching the spots across the gels.

Name	Gel	Filt	Landmrk
LabNoFiltNoLandmrk	Poured	No	No
LabFiltNoLandmrk	Poured	Yes	No
LabNoFiltLandmrk	Poured	No	Yes
PrecastNoFiltNoLandmrk	Bought	No	No
PrecastNoFiltLandmrk	Bought	No	Yes

pothesis testing, and the quest for improved test statistics that exploit the common information across the proteins (genes) [7, 8, 12, 5, 4].

As data from 2-D DIGE experiments exhibited similar characteristics to microarray datasets, methods developed by researchers in the microarray field were adapted to address statistical challenges in analyzing proteomic data from 2-D DIGE experiments. Section 3 describes the experiments used in this study and the proposed analysis methods. Section 4 contains the results, and Section 5 concludes with a summary.

## 3 Materials and methods

To ensure objective comparison of the methods, technical variability experiments were performed where the same biological sample was run on a number of different gels. In the absence of biological variation among the samples, the observed protein expression differences can be attributed to the technical variation in the 2-D DIGE process. The analysis methods can then be compared based on their ability to reduce the variability in the measurements corresponding to the same spot, and in their ability to reduce the number of proteins that appear to be differentially expressed. Human serum samples were used, to test the methods with the most complex, and difficult to analyze, proteome.

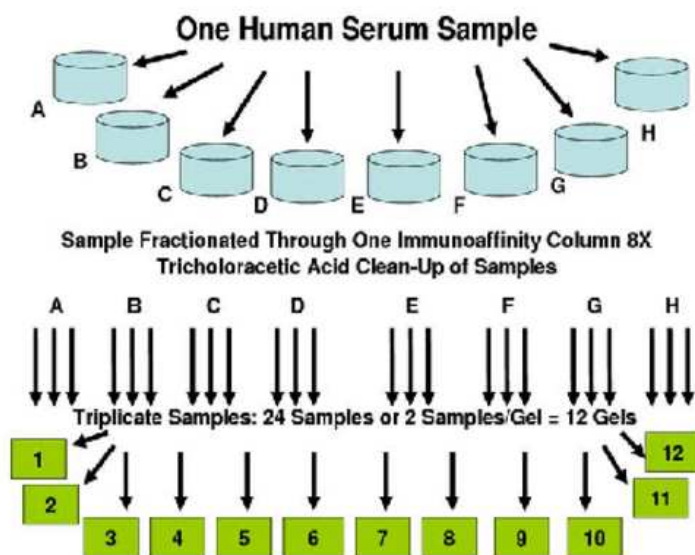


Figure 1: Sample processing workflow.

Five datasets were available for this study that differed on the type of gel used, and whether the spots were filtered or landmarked. Table 1 presents the details, along with the naming convention that will be used throughout this report. Three of the experiments involved gels poured in the lab, and two used Snap-A-Gel Precast gels manufactured by Jule Inc. (Milford, CT). To investigate the effect of eliminating spots that are likely to be dust particles or other artifacts, filtering was used in one dataset. In three experiments, spot matching was performed automatically across the gels, and two involved manual landmarking.

### 3.1 Sample preparation and gel processing

Blood was collected via venipuncture from a volunteer with informed consent under the Institutional Review Board approval from Lawrence Livermore National Laboratory. The blood was then split into 8 samples labeled A through H as shown in Fig. 1. After the sample processing steps (detailed in the remaining of this section), the 24 samples were arranged on 12 gels according to the

experimental design in Table 2.

Samples were processed via HPLC to remove the top 6 high-abundant proteins (albumin, IgG, antitrypsin, IgA, transferrin, and haptoglobin) using an Agilent Technologies HU6 Multiple Affinity column.  $40\mu L$  of each sample was combined with  $160\mu L$  Buffer A in a spin filter tube and filtered by spinning for 1 minute at 12000 rpm. Samples were loaded into a Shimadzu injector tube and placed into the sample injector rack. The AgilentHuman.met method was used along with the 8/11/04 Plasma Clean-up via FPLC/HPLC protocol.

$1000\mu L$  of the low-abundant protein fraction was cleaned-up using the 2/23/04 TCA + Amersham Wash Additive Protocol and was resuspended in a volume of  $75\mu L$ . The protein concentration of each cleaned-up sample was found using the ADV01 protein assay.  $50\mu g$  of each cleaned sample was labeled with 400pm of the appropriate CyDye for each analytical gel according to the experimental design in Table 2.

The labeled samples were separated by pI using Immobilized pH 3-10 NL DryStrips following the 2/23/04 pro-

Table 2: The experimental design.

Gel #	Cy3	Cy5	Cy2
1	A	F	pooled standard
2	G	B	pooled standard
3	E	H	pooled standard
4	D	A	pooled standard
5	F	D	pooled standard
6	C	G	pooled standard
7	E	B	pooled standard
8	A	C	pooled standard
9	H	D	pooled standard
10	C	F	pooled standard
11	B	H	pooled standard
12	G	E	pooled standard

tolcol for DIGE Labeling and First Dimension using the IPGphor II. The samples were alkylated and rehydrated using DTT followed by IAA following the 2/23/04 protocol for Second Dimension (SDS-PAGE). Samples were electrophoresed on Jule Snap-A-Gels (12.5D10ELOG) 12.5% acrylamide with a 38:1 acrylamide to bisacrylamide ratio. Second dimension electrophoresis was carried out in the Ettan Dalt Twelve Separation Unit following the 2/23/04 protocol for Second Dimension (SDS-PAGE). A deviation from the lab protocol was made for the SDS concentration in the 1x and 2x running buffers in order to follow the Jule’s suggested SDS concentration for Snap-A-Gels. The gels poured in the lab were prepared following standard 2-D DIGE protocol recommended by GE Healthcare.

The gel images were scanned on the Typhoon using the 2/23/04 Imaging DIGE Gels on the Typhoon protocol. The images were cropped using ImageQuant v5.2 to eliminate the edges of the gels and any major defects, while maximizing the number of spots available for analysis. By not cropping areas with minor defects the number of spots requiring manual verification increases, but so does the opportunity to find differentially expressed proteins. All gel images were cropped to be the same size.

Images were renamed to allow quicker references within DeCyder and processed using the DeCyder Batch Processor. The estimated number of spots to detect on

each gel was set to 2500, after manually exploring the best number of spots using the Differential In-gel Analysis (DIA) module of DeCyder. The master gel was automatically set to the gel with the greatest number of spots. If visual inspection indicated problems on the automatically selected master gel, the master gel was re-selected manually. Each sample was grouped for Biological Variation Analysis (BVA) according to the A - H sample names for the primary analysis. The original BVA files from the fully automated analysis were saved for future analysis.

Filtering to eliminate possible dust particles and other artifacts from the analysis was performed on the LabFiltNoLandmrk dataset (Table 1). Spots were eliminated from future analyses based on the following thresholds for four variables available in DeCyder: Slope > 1.1, Area < 100, Peak Height < 100, and Volume < 10000. For each spot, Slope refers to the gradient in the three-dimensional intensity view, Area is the number of pixels within the spot boundary, Peak Height is the maximum intensity value over the spot boundary, and Volume is the integrated intensity over the spot boundary.

The LabNoFiltLandmrk and the PouredNoFiltLandmrk (Table 1) datasets involved manual landmarking. Gels were landmarked to the master gel until all areas appeared to be matched correctly. When areas were found to be incorrectly matched additional landmarks were added and the gels were re-matched. Once the quality of matching for all gels was deemed sufficient analysis proceeded to determine the variation between samples.

### 3.2 Analysis using DeCyder

DeCyder version 5.01 was used for spot detection and matching across the gels. Both the Differential In-gel Analysis (DIA) and the Biological Variation Analysis (BVA) modules were used: the former to codetect and quantify the spots on a given gel in terms of ratios of the Cy3 and Cy5 sample volumes to the standard Cy2 volume, and the latter to match the spots and standardize the ratios across the gels accounting for differences in the Cy2 volumes on the gels.

For each spot on each gel in each set of experiments, DeCyder provided measurements on several variables. The analyses here were based on the spot volumes, normalized volumes, and standardized log abundance ratios defined below. Volume of a spot for a given dye is de-

defined as the fluorescent intensity of the corresponding dye integrated over the area of a spot. Normalized volume refers to the volume normalized across the three dyes and across the gels. Let  $VolCy2_{pg}$ ,  $VolCy3_{pg}$ , and  $VolCy5_{pg}$  denote the volumes of spot  $p$  on gel  $g$ , as measured by the three different dyes Cy2, Cy3, and Cy5, respectively. Similarly, let  $NVolCy2_{pg}$ ,  $NVolCy3_{pg}$ , and  $NVolCy5_{pg}$  denote the corresponding normalized volumes. To quantify the protein expression changes in a comparable manner across all the spots and gels, DeCyder forms the ratios of the normalized volumes of a spot labeled with the Cy3 or the Cy5 dye to the normalized volume of its corresponding internal standard labeled with Cy2. The resulting ratios are called the standardized abundances, and are denoted here by  $R32_{pg}$  and  $R52_{pg}$ :

$$\begin{cases} R32_{pg} = NVolCy3_{pg}/NVolCy2_{pg}, \\ R52_{pg} = NVolCy5_{pg}/NVolCy2_{pg}. \end{cases} \quad (1)$$

The analyses in DeCyder are based on the standardized log abundances of the proteins, defined as the  $\log_{10}$  of the standardized abundances, denoted here by  $SLA32_{pg}$  and  $SLA52_{pg}$ :

$$\begin{cases} SLA32_{pg} = \log_{10}(R32_{pg}), \\ SLA52_{pg} = \log_{10}(R52_{pg}). \end{cases} \quad (2)$$

Expression ratios corresponding to the standardized abundances in Eq. (1) are defined as

$$ER32_{pg} = \begin{cases} R32_{pg}, & \text{if } R32_{pg} \geq 1 \\ -1/R32_{pg}, & \text{otherwise} \end{cases}, \quad (3)$$

and

$$ER52_{pg} = \begin{cases} R52_{pg}, & \text{if } R52_{pg} \geq 1 \\ -1/R52_{pg}, & \text{otherwise} \end{cases}. \quad (4)$$

A spot's fold-change between two groups is calculated as the ratio of the average standardized abundances: if  $\bar{R}_1$  and  $\bar{R}_2$  denote the average standardized abundances corresponding to the two groups, the fold-change  $F$  is  $\bar{R}_1/\bar{R}_2$  if  $\bar{R}_1 \geq \bar{R}_2$ , and  $-\bar{R}_2/\bar{R}_1$  otherwise. A  $k$ -fold expression increase/decrease corresponds to a  $+k/-k$  value of  $F$ .

### 3.3 Additional statistical analyses

The variables that resulted from the analysis with DeCyder were exported and converted to text files with

the XML Toolbox. Additional statistical analyses were performed in the R open source computing environment (<http://www.r-project.org>). Specifically, the `marrayNorm` package from the Bioconductor project [5] was modified to accommodate the slightly different data structures characteristic to 2-D gels. In contrast with gene expression arrays that typically have only one or two colors, the gels have three separate measurements corresponding to the three dyes. Once the microarray algorithms were appropriately modified for the gel data, the normalization proceeded similar to the microarray normalization. Three different additional normalizations were performed, all based on the same idea but involving different variables. They are described in Sections 3.3.1, 3.3.2, and 3.3.3.

#### 3.3.1 Additional normalization based on the standardized log abundances

The first normalization method involved normalizing the standardized abundances from DeCyder. The standardized abundances in Eq. (1) were first transformed into the  $M - A$  space, where

$$\begin{cases} M_{pg}^{(1)} = \log_2(R52_{pg}/R32_{pg}), \\ A_{pg}^{(1)} = 1/2 \log_2(R52_{pg} \times R32_{pg}). \end{cases} \quad (5)$$

$A_{pg}^{(1)}$  measures the Average, and  $M_{pg}^{(1)}$  (Minus) the difference between the two intensity ratios  $R32_{pg}$  and  $R52_{pg}$ , on a log scale. The superscript  $(1)$  merely indicates that the values correspond to the first normalization type.

Under the assumption that the majority of the proteins were not differentially expressed between the two conditions (an assumption that is certainly satisfied under the current experimental design of using the same sample on each gel and dye), the plot of  $M_{pg}^{(1)}$  versus  $A_{pg}^{(1)}$  (MvA) for all the spots from a given gel should result in a random scatter around the zero-line with no trends. Observed systematic variations may be due to variations in the labeling efficiencies of the Cy3 and Cy5 dyes, as well as changing scanning settings, or gel effects. In microarrays, dye imbalances often vary according to the average spot intensity  $A$  [12]. Similar patterns were observed in the MvA plots of the 2-D DIGE datasets, therefore local intensity-dependent regression lines through the data were fitted using the `loessFit` function in R. Next, the  $M$  values were

replaced by the residuals from the fit which resulted in pattern-free MvA plots.

The second normalization step used boxplots for between-gel normalization. It involved comparing the ranges of the regression-corrected  $M$  values across the gels, and scaling them so that the middle 50 percent of the data on each gel spanned the same range.

Let  $\tilde{M}_{pg}^{(1)}$  and  $\tilde{A}_{pg}^{(1)}$  denote the corrected values after the MvA normalization within-gels and boxplot normalization between-gels. Next, the inverse transformation of Eq. (5) was used to transform  $\tilde{M}_{pg}^{(1)}$  and  $\tilde{A}_{pg}^{(1)}$  back to the original scale of the ratios, and thus obtain the normalized standardized abundances  $\tilde{R}32_{pg}^{(1)}$  and  $\tilde{R}52_{pg}^{(1)}$  corresponding to Eq. (1). The final step was to transform the normalized standardized abundances into the normalized standardized log abundances  $\tilde{S}LA32_{pg}^{(1)}$  and  $\tilde{S}LA52_{pg}^{(1)}$  with the logarithm transformation in Eq. (2). The standardized log abundances from DeCyder were thus further normalized.

### 3.3.2 Additional normalization based on the normalized volumes

The second normalization method involved two separate MvA and boxplot normalizations of the type described in Section 3.3.1. First, the normalized volumes (provided by DeCyder) measured with the Cy3 dye were additionally normalized to the normalized volumes measured with the pooled standard Cy2. If

$$\begin{cases} M_{pg}^{(2)} = \log_2(NVolCy3_{pg}/NVolCy2_{pg}) \\ A_{pg}^{(2)} = 1/2 \log_2(NVolCy3_{pg} \times NVolCy2_{pg}) \end{cases} \quad (6)$$

denote the normalized Cy2 and Cy3 volumes transformed to the  $M - A$  space, let  $\tilde{M}_{pg}^{(2)}$  and  $\tilde{A}_{pg}^{(2)}$  denote their corrected values after the MvA normalization within-gels and boxplot normalization between-gels. The  $\tilde{M}_{pg}^{(2)}$  and  $\tilde{A}_{pg}^{(2)}$  values were next transformed back to the original scale of the normalized volumes using the inverse transformation in Eq. (6), and thus the additionally normalized normalized volumes  $\tilde{NVolCy3}_{pg}^{(2)}$  and  $\tilde{NVolCy2}_{pg}^{(2)}$  were obtained. Next, the ratios

$$\tilde{R}32_{pg}^{(2)} = \tilde{NVolCy3}_{pg}^{(2)} / \tilde{NVolCy2}_{pg}^{(2)} \quad (7)$$

were formed which corresponded to the normalized standardized abundances of the Cy3 to Cy2 ratios in Eq. (1).

A similar procedure was performed to additionally normalize the normalized Cy5 volumes to the normalized Cy2 volumes, and obtain

$$\tilde{R}52_{pg}^{(2)} = \tilde{NVolCy5}_{pg}^{(2)} / \tilde{NVolCy2}_{pg}^{(2)}, \quad (8)$$

the normalized standardized abundances. The normalization steps in the previous paragraph were executed, starting with  $NVolCy5_{pg}$  instead of  $NVolCy3_{pg}$  in Eq. (6).

The final step was to transform the normalized standardized abundances in Eqs. (7) and (8) into the normalized standardized log abundances  $\tilde{S}LA32_{pg}^{(2)}$  and  $\tilde{S}LA52_{pg}^{(2)}$  by applying the logarithm transformation.

### 3.3.3 Additional normalization based on the original volumes

The third normalization method was identical to the second procedure in Section 3.3.2 – the only difference was using the original volume measurements from DeCyder rather than the normalized volumes. The third set of additionally normalized standardized log abundances  $\tilde{S}LA32_{pg}^{(3)}$  and  $\tilde{S}LA52_{pg}^{(3)}$  was obtained by following the steps in Section 3.3.2 after substituting  $VolCy2_{pg}$ ,  $VolCy3_{pg}$ , and  $VolCy5_{pg}$  for  $NVolCy2_{pg}$ ,  $NVolCy3_{pg}$ , and  $NVolCy5_{pg}$ .

### 3.3.4 Adjusting the significance levels in multiple hypothesis tests

Another challenge in the analysis of 2-D DIGE data that is shared with the microarray data analysis community is the massive multiple hypothesis problem [11]. To investigate whether a protein is differentially expressed across the conditions in a study, generally a statistical model is fit at every spot in the experiment, and a hypothesis test is performed to assess the significance of the observed test statistic. If the observed significance level, or p-value, is less than a specified significance level value (say  $\alpha = 0.05$ ), then a statistically significant result is declared. Regardless of the data used and the testing procedure employed, the resulting p-values need to be adjusted because numerous tests are performed simultaneously. The unadjusted p-values that result from the individual tests applied separately at each spot are too optimistic. At the  $\alpha = 0.05$  significance level, one in every



twenty tests is expected to result in a p-value less than  $\alpha$  just by chance. As the number of tests increases, so does the number of false positives.

Several adjustment methods have been proposed in the literature. The simplest one is the Bonferroni correction, which multiplies the unadjusted p-values by the total number of tests performed. A less stringent, but more practical approach for the present case is the false discovery rate method of [3]. Let  $R$  denote the total number of rejected hypotheses, and  $V$  the number of falsely rejected hypotheses, out from the total number of simultaneous tests. Then, the realized False Discovery Rate (FDR) is defined as  $V/R$ , for  $R > 0$ , and 0 otherwise. Since  $V$  is unobserved, [3] developed a sequential p-value procedure that controls the *expected* value of the FDR,  $E(FDR)$ , under the assumption that the test statistics are independent. The resulting process controls  $E(FDR)$  at the fixed level  $\alpha$  for any joint distribution of the p-values. Although the independence assumption is not always satisfied, the FDR method is often used because of its simplicity. Since its results are preferable over the unadjusted p-values, here the FDR procedure in  $R$  was used.

## 4 Results and discussion

### 4.1 Comparison of the datasets

Fig. 2 displays the spot matching results for the three experiments using gels poured in the lab. Fig. 3 shows the corresponding results with the precast gels. Table 3 presents the corresponding bin counts and cumulative matching percentages for the data in Figs. 2 and 3.

Notice the relatively high number of spots in the leftmost bins of the histograms in Figs. 2 and 3 (equivalently, in the rightmost columns of Table 3): those spots were matched on at most two gels, and include dust particles and other artifacts identified on the master gel only. In the case of the gels poured in the lab presented in Fig. 2, filtering shown in panel (b) decreased the percentage of spots that were matched on more than 8 gels from the percentage observed in the original dataset in panel (a). The filtering in panel (c) increased the percent of spots matched on over 8 gels. Filtering also improved considerably the quality of the matching for the precast gels: note the higher percentage of spots matched on at least 8 gels

in Fig. 3(b) compared to Fig. 3(a), quantified in the 5th and 4th rows of Table 3, respectively.

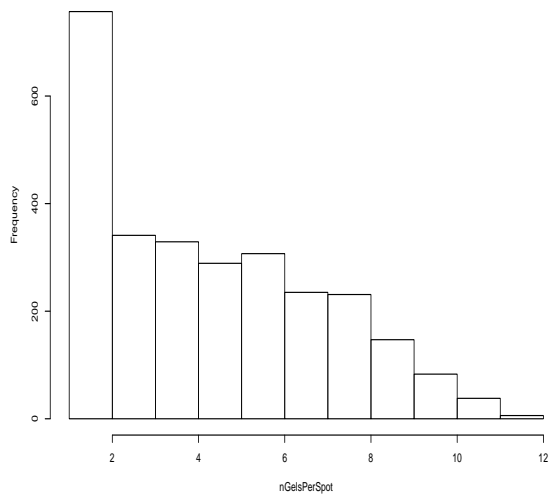
Overall, the results indicate that filtering did not improve the matching compared to the baseline of no filtering and no landmarking for the gels poured in the lab. In contrast, landmarking improved the accuracy of the matching, both for the gels poured in-house and the precast gels. Compared to the in-house gels, the precast gels led to more accurate spot matching.

Tables 4 through 13 present summary statistics for the spots-wise standard deviations of the standardized log abundances for the five datasets. First, for each spot in a dataset, the standard deviation of its standardized log abundance in Eq. (2) over the gels in the dataset was calculated. Next, spots were grouped based on the matching results over the gels. The tables present summary statistics of the distribution of the resulting standard deviations over the subsets. As expected, the better the match was in a subset of the spots, the less variation was observed among the corresponding spots. Overall, there was less variability in the precast gels than in the in-house gels in each of the spot subsets considered.

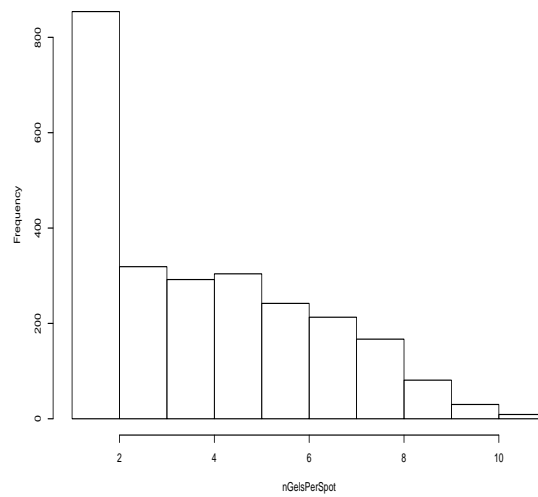
Fig. 4 displays visually the trend of decreasing standard deviation with increased matching accuracy: it shows boxplots of the standard deviations of the standardized log abundances for the PrecastNoFiltLandmrk dataset, split by subsets of spots based on the matching. The values correspond to the combined *SLA32* and *SLA52* data in Table 13.

To ensure high confidence in the results, spots poorly matched may be discarded from future statistical analysis. Requiring that a spot be matched on at least 70% of the gels (8 out of 12 gels here) seems like a reasonable criterion. The corresponding reduction in the number of spots reflects the current state of the art in spot detection and matching in complex proteomic samples. In real biological experiments, the biological variation among the samples will further complicate the spot detection and matching, and will likely be of inferior quality.

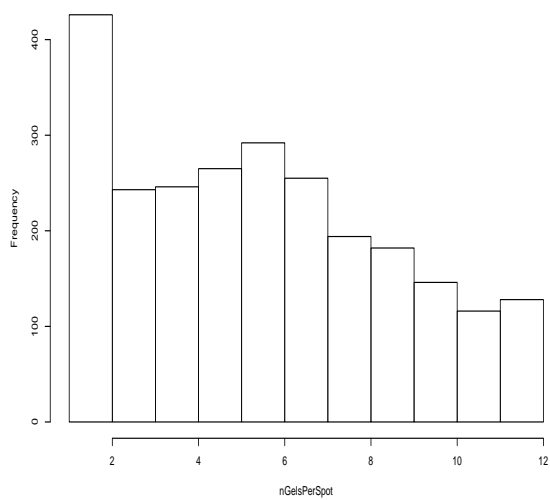
Fig. 5 reveals an expected relationship between the magnitude of the raw spot volumes and the quality of the matching: the larger the volume the better the match. Similar trends were observed for the other dyes and gels. Considering two subsets of the spots: 1) those matched on fewer than 8 gels and 2) those matched on at least 8 gels, the interquartile ranges for the  $\log_{10}(VolCy2)$



(a)



(b)



(c)

Figure 2: Histogram of the number of gels a spot was matched on for the gels poured in the lab: (a) No filtering and no landmarking (LabNoFiltNoLandmrk), (b) Filtering and no landmarking (LabFiltNoLandmrk), (c) No filtering and landmarking (LabNoFiltLandmrk).

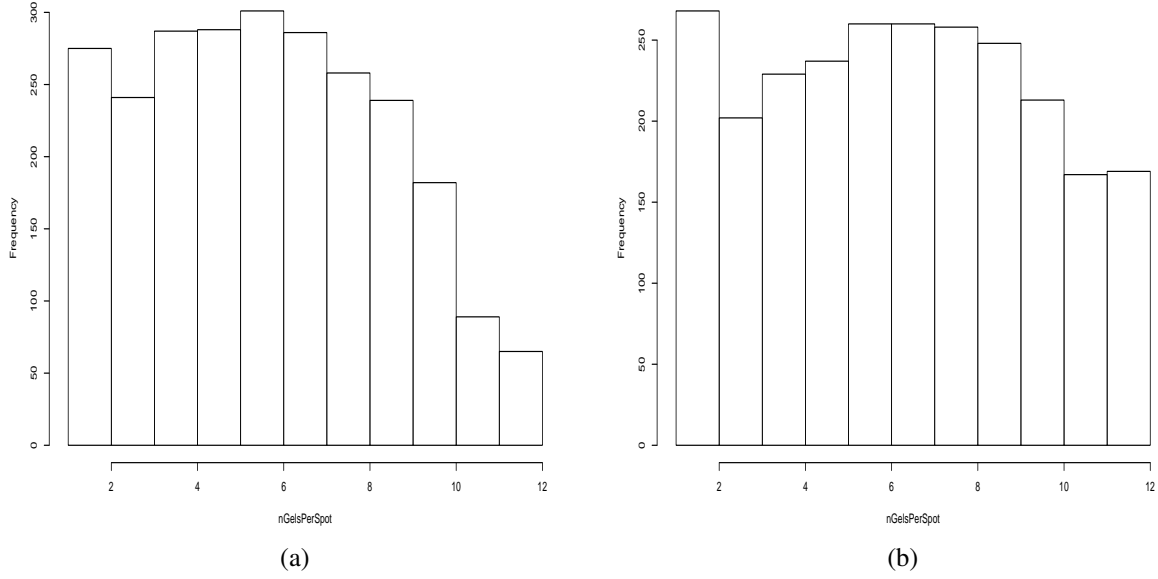


Figure 3: Histogram of the number of gels a spot was matched on for the precast gels: (a) No filtering and no landmarking (PrecastNoFiltNoLandmrk), (b) No filtering and landmarking (PrecastNoFiltLandmrk).

Table 3: The distribution of the spots in the five datasets based on their matching status across the twelve gels. The columns indicate the number of gels a spot was matched on, in decreasing order from 12 to 1-2. The five rows separated by the lines correspond to the five datasets. For each dataset, the first row in the column with header  $j$  displays the number of spots matched on exactly  $j$  gels, for  $j = 12, 11, \dots, 1 - 2$ . The second row presents the cumulative percentage of spots from the corresponding dataset matched on at least  $j$  gels.

Data	12	11	10	9	8	7	6	5	4	3	1-2
Lab	6	38	83	147	231	235	307	289	329	341	757
NoFiltNoLandmrk	0.22	1.59	4.60	9.92	18.28	26.78	37.89	48.35	60.26	72.60	100
Lab	0	9	30	81	164	207	236	284	270	291	822
FiltNoLandmrk	0	0.38	1.63	5.01	11.86	20.51	30.37	42.23	53.51	65.66	100
Lab	128	116	146	182	194	255	292	265	246	243	426
NoFiltLandmrk	5.13	9.79	15.64	22.94	30.73	49.95	52.67	63.30	73.16	82.91	100
Precast	65	89	182	239	258	286	301	288	287	241	275
NoFiltNoLandmrk	2.59	6.13	13.38	22.90	33.17	44.56	56.55	68.02	79.45	89.05	100
Precast	169	167	213	248	258	260	260	237	229	202	268
NoFiltLandmrk	6.73	13.38	21.86	31.74	42.02	52.37	62.72	71.16	81.28	89.33	100

Table 4: LabNoFiltNoLandmrk dataset: summary statistics for the standard deviations of the standardized log abundances for certain subsets of the spots. MatchStatus indicates the criteria of selecting the subsets, mainly the number of gels the spots were matched on. The disjoint subsets include spots matched on all 12 gels, and spots matched on either  $2*j$  or  $(2*j+1)$  gels for  $j = 5, \dots, 2$ . The NoSpots column indicates the number of spots that belonged to the corresponding spot subset. The Data rows indicate which dyes the standardized log abundance ratios corresponded to: only *SLA32*, only *SLA52*, or combined *SLA32* and *SLA52* from Eq. (2).

Data	MatchStatus	NoSpots	Min	1stQu	Med	Mean	3rdQu	Max
<i>SLA32</i>	12	6	0.0470	0.0635	0.0706	0.0710	0.0783	0.0956
	10_11	121	0.0235	0.0452	0.0623	0.0661	0.0811	0.1459
	8_9	378	0.0151	0.0480	0.0623	0.0673	0.0787	0.2273
	6_7	542	0.0118	0.0506	0.0716	0.0862	0.1030	0.4492
	4_5	618	0.0140	0.0562	0.0944	0.1183	0.1607	0.6160
	2_3	742	0.0002	0.0570	0.1183	0.1509	0.2050	0.8842
<i>SLA52</i>	12	6	0.0344	0.0587	0.0662	0.0879	0.0697	0.2336
	10_11	121	0.0219	0.0433	0.0578	0.0704	0.0832	0.2554
	8_9	378	0.0140	0.0431	0.0624	0.0767	0.0892	0.4051
	6_7	542	0.0138	0.0511	0.0753	0.1152	0.1249	0.6754
	4_5	618	0.0081	0.0639	0.1253	0.1929	0.2969	0.9457
	2_3	742	0.0000	0.0833	0.2217	0.2948	0.4086	1.8840
<i>SLA32</i> and <i>SLA52</i>	12	6	0.0423	0.0688	0.0713	0.0841	0.0797	0.1695
	10_11	121	0.0289	0.0515	0.0649	0.0708	0.0801	0.1960
	8_9	378	0.0284	0.0509	0.0651	0.0756	0.0874	0.3186
	6_7	542	0.0209	0.0550	0.0763	0.1150	0.1233	0.6611
	4_5	618	0.0218	0.0642	0.1229	0.2018	0.3139	0.7369
	2_3	742	0.0135	0.1472	0.3859	0.3807	0.5661	1.2850

Table 5: LabNoFiltNoLandmrk dataset: summary statistics for the standard deviations of the standardized log abundances for certain subsets of the spots. MatchStatus indicates the criteria of selecting the subsets, mainly the number of gels the spots were matched on. The subsets include spots matched on at least  $2 * j$  gels for  $j = 5, \dots, 1$ . The NoSpots column indicates the number of spots that belonged to the corresponding spot subset. The Data rows indicate which dyes the standardized log abundance ratios corresponded to: only *SLA32*, only *SLA52*, or combined *SLA32* and *SLA52* from Eq. (2).

Data	MatchStatus	NoSpots	Min	1stQu	Med	Mean	3rdQu	Max
<i>SLA32</i>	10+	127	0.0235	0.0461	0.0624	0.0663	0.0809	0.1459
	8+	505	0.0151	0.0476	0.0624	0.0670	0.0795	0.2273
	6+	1047	0.0118	0.0487	0.0668	0.0769	0.0889	0.4492
	4+	1665	0.0118	0.0502	0.0726	0.0923	0.1110	0.6160
	2+	2407	0.0002	0.0512	0.0788	0.1104	0.1349	0.8842
<i>SLA52</i>	10+	127	0.0219	0.0449	0.0598	0.0712	0.0827	0.2554
	8+	505	0.0140	0.0432	0.0618	0.0753	0.0858	0.4051
	6+	1047	0.0138	0.0477	0.0673	0.0959	0.1029	0.6754
	4+	1665	0.0081	0.0507	0.0779	0.1319	0.1554	0.9457
	2+	2407	0.0000	0.0546	0.0948	0.1821	0.2501	1.8840
<i>SLA32</i> and <i>SLA52</i>	10+	127	0.0289	0.0516	0.0668	0.0714	0.0805	0.1960
	8+	505	0.0284	0.0512	0.0653	0.0746	0.0854	0.3186
	6+	1047	0.0209	0.0526	0.0698	0.0955	0.0982	0.6611
	4+	1665	0.0209	0.0550	0.0778	0.1349	0.1478	0.7369
	2+	2407	0.0135	0.0605	0.1017	0.2107	0.3227	1.2850

Table 6: LabFiltNoLandmrk dataset: summary statistics for the standard deviations of the standardized log abundances for certain subsets of the spots. Caption as in Table 4.

Data	MatchStatus	NoSpots	Min	1stQu	Med	Mean	3rdQu	Max
<i>SLA32</i>	12	0	NA	NA	NA	NA	NA	NA
	10_11	39	0.0224	0.0443	0.0595	0.0666	0.0909	0.1230
	8_9	245	0.0185	0.0472	0.0617	0.0679	0.0812	0.2009
	6_7	443	0.0138	0.0438	0.0598	0.0678	0.0824	0.3154
	4_5	554	0.0057	0.0439	0.0683	0.0831	0.1014	0.4700
	2_3	657	0.0001	0.0434	0.0781	0.1102	0.1437	0.7815
<i>SLA52</i>	12	0	NA	NA	NA	NA	NA	NA
	10_11	39	0.0280	0.0538	0.0735	0.0789	0.0930	0.1943
	8_9	245	0.0222	0.0471	0.0604	0.0729	0.0849	0.2982
	6_7	443	0.0095	0.0452	0.0625	0.0762	0.0848	0.4502
	4_5	554	0.0084	0.0487	0.0773	0.1101	0.1219	0.6919
	2_3	657	0.0001	0.0630	0.1462	0.2591	0.3382	2.0300
<i>SLA32</i> and <i>SLA52</i>	12	0	NA	NA	NA	NA	NA	NA
	10_11	39	0.0266	0.0561	0.0676	0.0737	0.0915	0.1571
	8_9	245	0.0293	0.0502	0.0605	0.0717	0.0821	0.2271
	6_7	443	0.0218	0.0496	0.0635	0.0748	0.0859	0.3540
	4_5	554	0.0180	0.0520	0.0730	0.1044	0.1210	0.5528
	2_3	657	0.0092	0.0777	0.1720	0.2654	0.4023	1.3380

Table 7: LabFiltNoLandmrk dataset: summary statistics for the standard deviations of the standardized log abundances for certain subsets of the spots. Caption as in Table 5.

Data	MatchStatus	NoSpots	Min	1stQu	Med	Mean	3rdQu	Max
<i>SLA32</i>	10+	39	0.0224	0.0443	0.0595	0.0666	0.0909	0.1230
	8+	284	0.0185	0.0466	0.0615	0.0677	0.0819	0.2009
	6+	727	0.0138	0.0453	0.0603	0.0678	0.0824	0.3154
	4+	1281	0.0057	0.0444	0.0632	0.0744	0.0890	0.4700
	2+	1938	0.0001	0.0444	0.0670	0.0865	0.1013	0.7815
<i>SLA52</i>	10+	39	0.0280	0.0538	0.0735	0.0789	0.0930	0.1943
	8+	284	0.0222	0.0480	0.0620	0.0738	0.0870	0.2982
	6+	727	0.0095	0.0467	0.0624	0.0752	0.0860	0.4502
	4+	1281	0.0084	0.0471	0.0671	0.0903	0.0978	0.6919
	2+	1938	0.0001	0.0496	0.0779	0.1475	0.1473	2.0300
<i>SLA32</i> and <i>SLA52</i>	10+	39	0.0266	0.0561	0.0676	0.0737	0.0915	0.1571
	8+	284	0.0266	0.0506	0.0620	0.0720	0.0823	0.2271
	6+	727	0.0218	0.0502	0.0633	0.0737	0.0841	0.3540
	4+	1281	0.0180	0.0506	0.0669	0.0870	0.0952	0.5528
	2+	1938	0.0092	0.0544	0.0781	0.1474	0.1590	1.3380

Table 8: LabNoFiltLandmrk dataset: summary statistics for the standard deviations of the standardized log abundances for certain subsets of the spots. Caption as in Table 4.

Data	MatchStatus	NoSpots	Min	1stQu	Med	Mean	3rdQu	Max
<i>SLA32</i>	12	128	0.0130	0.0292	0.0385	0.0467	0.0549	0.1737
	10_11	262	0.0163	0.0358	0.0498	0.0570	0.0692	0.1799
	8_9	376	0.0130	0.0435	0.0588	0.0692	0.0788	0.3452
	6_7	547	0.0167	0.0496	0.0654	0.0812	0.0938	0.5712
	4_5	511	0.0056	0.0483	0.0796	0.1139	0.1418	0.6667
	2_3	503	0.0002	0.0494	0.0989	0.1501	0.2136	0.7539
<i>SLA52</i>	12	128	0.0125	0.0251	0.0320	0.0422	0.0494	0.1400
	10_11	262	0.0145	0.0316	0.0465	0.0576	0.0684	0.2823
	8_9	376	0.0132	0.0366	0.0540	0.0707	0.0827	0.4103
	6_7	547	0.0090	0.0414	0.0640	0.0945	0.1048	0.8615
	4_5	511	0.0028	0.0499	0.0866	0.1586	0.2106	0.9313
	2_3	503	0.0003	0.0697	0.1749	0.2418	0.3621	1.3410
<i>SLA32</i> and <i>SLA52</i>	12	128	0.0188	0.0348	0.0418	0.0494	0.0557	0.1595
	10_11	262	0.0174	0.0401	0.0536	0.0627	0.0744	0.2122
	8_9	376	0.0183	0.0452	0.0613	0.0758	0.0879	0.3724
	6_7	547	0.0164	0.0486	0.0681	0.0975	0.1094	0.7089
	4_5	511	0.0093	0.0555	0.0893	0.1490	0.2025	0.7355
	2_3	503	0.0069	0.0966	0.2196	0.2521	0.3681	0.9069

Table 9: LabNoFiltLandmrk dataset: summary statistics for the standard deviations of the standardized log abundances for certain subsets of the spots. Caption as in Table 5.

Data	MatchStatus	NoSpots	Min	1stQu	Med	Mean	3rdQu	Max
<i>SLA32</i>	10+	390	0.0130	0.0326	0.0454	0.0537	0.0635	0.1799
	8+	766	0.0130	0.0371	0.0527	0.0613	0.0725	0.3452
	6+	1313	0.0130	0.0400	0.0569	0.0696	0.0813	0.5712
	4+	1824	0.0056	0.0419	0.0610	0.0820	0.0932	0.6667
	2+	2327	0.0002	0.0428	0.0646	0.0967	0.1085	0.7539
<i>SLA52</i>	10+	390	0.0125	0.0288	0.0410	0.0525	0.0638	0.2823
	8+	766	0.0125	0.0317	0.0474	0.0615	0.0742	0.4103
	6+	1313	0.0090	0.0340	0.0527	0.0752	0.0844	0.8615
	4+	1824	0.0028	0.0370	0.0596	0.0986	0.1040	0.9313
	2+	2327	0.0003	0.0398	0.0679	0.1295	0.1481	1.3410
<i>SLA32</i> and <i>SLA52</i>	10+	390	0.0174	0.0374	0.0480	0.0584	0.0680	0.2122
	8+	766	0.0174	0.0404	0.0552	0.0669	0.0770	0.3724
	6+	1313	0.0164	0.0432	0.0603	0.0796	0.0884	0.7089
	4+	1824	0.0093	0.0459	0.0649	0.0991	0.1083	0.7355
	2+	2327	0.0069	0.0478	0.0743	0.1321	0.1595	0.9069

Table 10: PrecastNoFiltNoLandmrk dataset: summary statistics for the standard deviations of the standardized log abundances for certain subsets of the spots. Caption as in Table 4.

Data	MatchStatus	NoSpots	Min	1stQu	Med	Mean	3rdQu	Max
<i>SLA32</i>	12	65	0.0188	0.0321	0.0384	0.0445	0.0506	0.0908
	10_11	271	0.0154	0.0310	0.0392	0.0453	0.0519	0.3571
	8_9	497	0.0112	0.0304	0.0433	0.0520	0.0617	0.4767
	6_7	587	0.0065	0.0350	0.0485	0.0624	0.0736	0.5772
	4_5	575	0.0050	0.0373	0.0581	0.0899	0.0999	0.7215
	2_3	423	0.0000	0.0321	0.0642	0.1126	0.1242	0.9734
<i>SLA52</i>	12	65	0.0176	0.0298	0.0407	0.0450	0.0542	0.1151
	10_11	271	0.0141	0.0316	0.0397	0.0450	0.0517	0.3045
	8_9	497	0.0086	0.0304	0.0441	0.0538	0.0598	0.5264
	6_7	587	0.0089	0.0340	0.0472	0.0642	0.0714	0.6332
	4_5	575	0.0047	0.0335	0.0545	0.0919	0.0966	0.8518
	2_3	423	0.0004	0.0267	0.0568	0.1089	0.1166	0.9695
<i>SLA32</i> and <i>SLA52</i>	12	65	0.0213	0.0322	0.0448	0.0460	0.0534	0.0988
	10_11	271	0.0207	0.0343	0.0423	0.0473	0.0529	0.3230
	8_9	497	0.0157	0.0334	0.0462	0.0574	0.0652	0.4665
	6_7	587	0.0137	0.0369	0.0509	0.0710	0.0798	0.5791
	4_5	575	0.0107	0.0422	0.0686	0.1066	0.1322	0.6963
	2_3	423	0.0036	0.0457	0.0787	0.1276	0.1471	0.7886

Table 11: PrecastNoFiltNoLandmrk dataset: summary statistics for the standard deviations of the standardized log abundances for certain subsets of the spots. Caption as in Table 5.

Data	MatchStatus	NoSpots	Min	1stQu	Med	Mean	3rdQu	Max
<i>SLA32</i>	10+	336	0.0154	0.0313	0.0391	0.0451	0.0519	0.3571
	8+	833	0.0112	0.0308	0.0414	0.0492	0.0574	0.4767
	6+	1420	0.0065	0.0318	0.0440	0.0547	0.0634	0.5772
	4+	1995	0.0050	0.0329	0.0472	0.0648	0.0710	0.7215
	2+	2418	0.0000	0.0328	0.0486	0.0732	0.0769	0.9734
<i>SLA52</i>	10+	336	0.0141	0.0314	0.0398	0.0450	0.0519	0.3045
	8+	833	0.0086	0.0310	0.0418	0.0502	0.0571	0.5264
	6+	1420	0.0086	0.0316	0.0439	0.0560	0.0625	0.6332
	4+	1995	0.0047	0.0320	0.0458	0.0664	0.0698	0.8518
	2+	2418	0.0004	0.0313	0.0467	0.0738	0.0748	0.9695
<i>SLA32</i> and <i>SLA52</i>	10+	336	0.0207	0.0336	0.0424	0.0471	0.0529	0.3230
	8+	833	0.0157	0.0335	0.0445	0.0533	0.0593	0.4665
	6+	1420	0.0137	0.0350	0.0468	0.0606	0.0674	0.5791
	4+	1995	0.0107	0.0362	0.0498	0.0738	0.0787	0.6963
	2+	2418	0.0036	0.0370	0.0524	0.0833	0.0877	0.7886



Table 12: PrecastNoFiltLandmrk dataset: summary statistics for the standard deviations of the standardized log abundances for certain subsets of the spots. Caption as in Table 4.

Data	MatchStatus	NoSpots	Min	1stQu	Med	Mean	3rdQu	Max
<i>SLA32</i>	12	169	0.0164	0.0310	0.0389	0.0431	0.0502	0.1551
	10_11	380	0.0070	0.0294	0.0394	0.0460	0.0562	0.2354
	8_9	506	0.0113	0.0308	0.0423	0.0531	0.0612	0.4178
	6_7	520	0.0090	0.0330	0.0476	0.0638	0.0722	0.5959
	4_5	466	0.0066	0.0344	0.0556	0.0816	0.0929	0.6707
	2_3	373	0.0011	0.0272	0.0571	0.1006	0.1057	0.7920
<i>SLA52</i>	12	169	0.0166	0.0315	0.0400	0.0452	0.0512	0.1533
	10_11	380	0.0114	0.0294	0.0385	0.0467	0.0541	0.2605
	8_9	506	0.0068	0.0284	0.0391	0.0490	0.0586	0.4423
	6_7	520	0.0087	0.0302	0.0424	0.0598	0.0661	0.6620
	4_5	466	0.0068	0.0307	0.0488	0.0782	0.0811	0.7563
	2_3	373	0.0004	0.0260	0.0519	0.1037	0.1035	1.0570
<i>SLA32</i> and <i>SLA52</i>	12	169	0.0192	0.0342	0.0404	0.0465	0.0539	0.1831
	10_11	380	0.0167	0.0334	0.0424	0.0511	0.0568	0.2010
	8_9	506	0.0144	0.0324	0.0441	0.0582	0.0677	0.4183
	6_7	520	0.0137	0.0356	0.0496	0.0718	0.0822	0.6022
	4_5	466	0.0182	0.0406	0.0611	0.0983	0.1249	0.6678
	2_3	373	0.0096	0.0397	0.0642	0.1239	0.1446	0.7889

Table 13: PrecastNoFiltLandmrk dataset: summary statistics for the standard deviations of the standardized log abundances for certain subsets of the spots. Caption as in Table 5.

Data	MatchStatus	NoSpots	Min	1stQu	Med	Mean	3rdQu	Max
<i>SLA32</i>	10+	549	0.0070	0.0298	0.0393	0.0451	0.0533	0.2354
	8+	1055	0.0070	0.0303	0.0410	0.0489	0.0571	0.4178
	6+	1575	0.0070	0.0308	0.0429	0.0539	0.0608	0.5959
	4+	2041	0.0066	0.0316	0.0448	0.0602	0.0661	0.6707
	2+	2414	0.0011	0.0312	0.0456	0.0664	0.0704	0.7920
<i>SLA52</i>	10+	549	0.0114	0.0304	0.0392	0.0462	0.0534	0.2605
	8+	1055	0.0068	0.0294	0.0392	0.0476	0.0562	0.4423
	6+	1575	0.0068	0.0296	0.0402	0.0516	0.0586	0.6620
	4+	2041	0.0068	0.0298	0.0418	0.0577	0.0622	0.7563
	2+	2414	0.0004	0.0294	0.0427	0.0648	0.0652	1.0570
<i>SLA32</i> and <i>SLA52</i>	10+	549	0.0167	0.0336	0.0422	0.0497	0.0558	0.2010
	8+	1055	0.0144	0.0330	0.0427	0.0538	0.0597	0.4183
	6+	1575	0.0137	0.0338	0.0446	0.0597	0.0665	0.6022
	4+	2041	0.0137	0.0346	0.0471	0.0685	0.0750	0.6678
	2+	2414	0.0096	0.0350	0.0489	0.0771	0.0827	0.7889

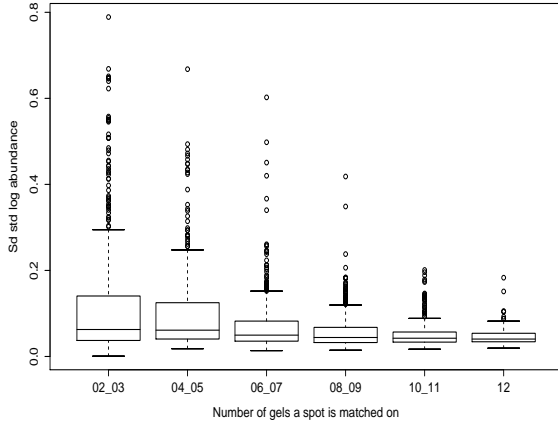


Figure 4: The distribution of the spot-wise standard deviations of the combined standardized log abundances for the PrecastNoFiltLandmrk dataset, split by the matching status of the spots: 02\_03 includes spots matched on 2 or 3 gels, 12 includes spots matched on all 12 gels.

corresponding to the two sets are 1) [4.089, 4.842] and 2) [4.843, 5.786]. Therefore, an empirical rule to select spots to include in the analysis in future experiments with human serum is to consider only spots with normalized volumes over 70,000 ( $10^{4.842}$ ).

## 4.2 Effect of additional normalizations

This section presents the results of additional statistical normalizations performed on the data that resulted from DeCyder. Only results with the PrecastNoFiltLandmrk experiment are reported, which was the best of the five sets considered. The corresponding analyses with the other four datasets yielded similar results, and are omitted for brevity. Their details are available upon request.

Fig. 6 displays the  $M - A$  plots for four gels, obtained through the transformation in Eq. (5) with the first normalization method in Section 3.3.1. Fig. 7 shows the corresponding  $\tilde{M} - \tilde{A}$  plots after correcting the intensity-dependent bias with a local regression fit. Fig. 8 shows the boxplots of the data before and after the normalization in Section 3.3.1 based on further correcting the standardized

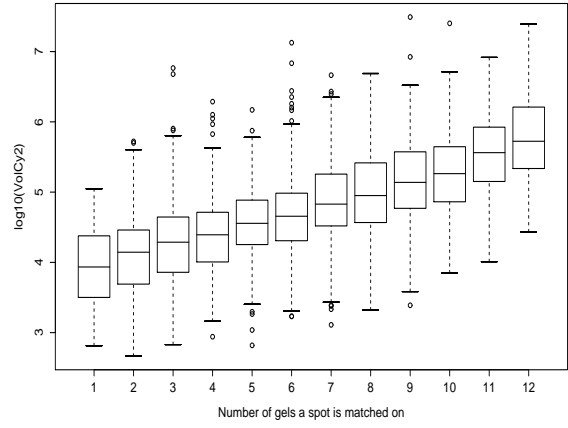


Figure 5: Boxplot of the log10 of the original Cy2 volumes of the spots on the first gel in the PrecastNoFiltLandmrk dataset, as a function of the number of gels a spot was matched on.

log abundances from DeCyder. The two alternative normalization methods (based on the normalized volumes in Section 3.3.2, and on the volumes in Section 3.3.3, respectively) yielded qualitatively similar results.

Figs. 9 and 10 display the densities and boxplots of the original and the three normalized standardized log abundances separately for the twelve gels and two color ratios. The apparent differences between the red and green density estimates on the twelve gels in the first panel of Fig. 9 are mitigated by all three normalization methods, and especially by the last two methods explained in Sections 3.3.2 and 3.3.3.

Based on the distribution of the expression ratios on the gels, prediction intervals for the expression ratios on a future gel can be constructed. Given the mean  $m_{2.5}$  and sample variance  $s_{2.5}^2$  of the twelve 2.5 percentile values of the expression ratios  $ER_{32}$  or  $ER_{52}$  from Eqs. (3) or (4) corresponding to the twelve gels, an approximate 95% prediction limit for the 2.5 percentile on a future gel can be obtained as follows:  $m_{2.5} \pm t_{\alpha/2, n-1} s_{2.5} \sqrt{1 + 1/n}$ , where  $n = 12$  is the number of gels,  $\alpha = 0.05$  and  $t_{\alpha/2, n-1}$  is the  $100(1 - \alpha/2)$  percentile of the  $t$ -distribution with  $(n - 1)$  degrees of freedom. Similar

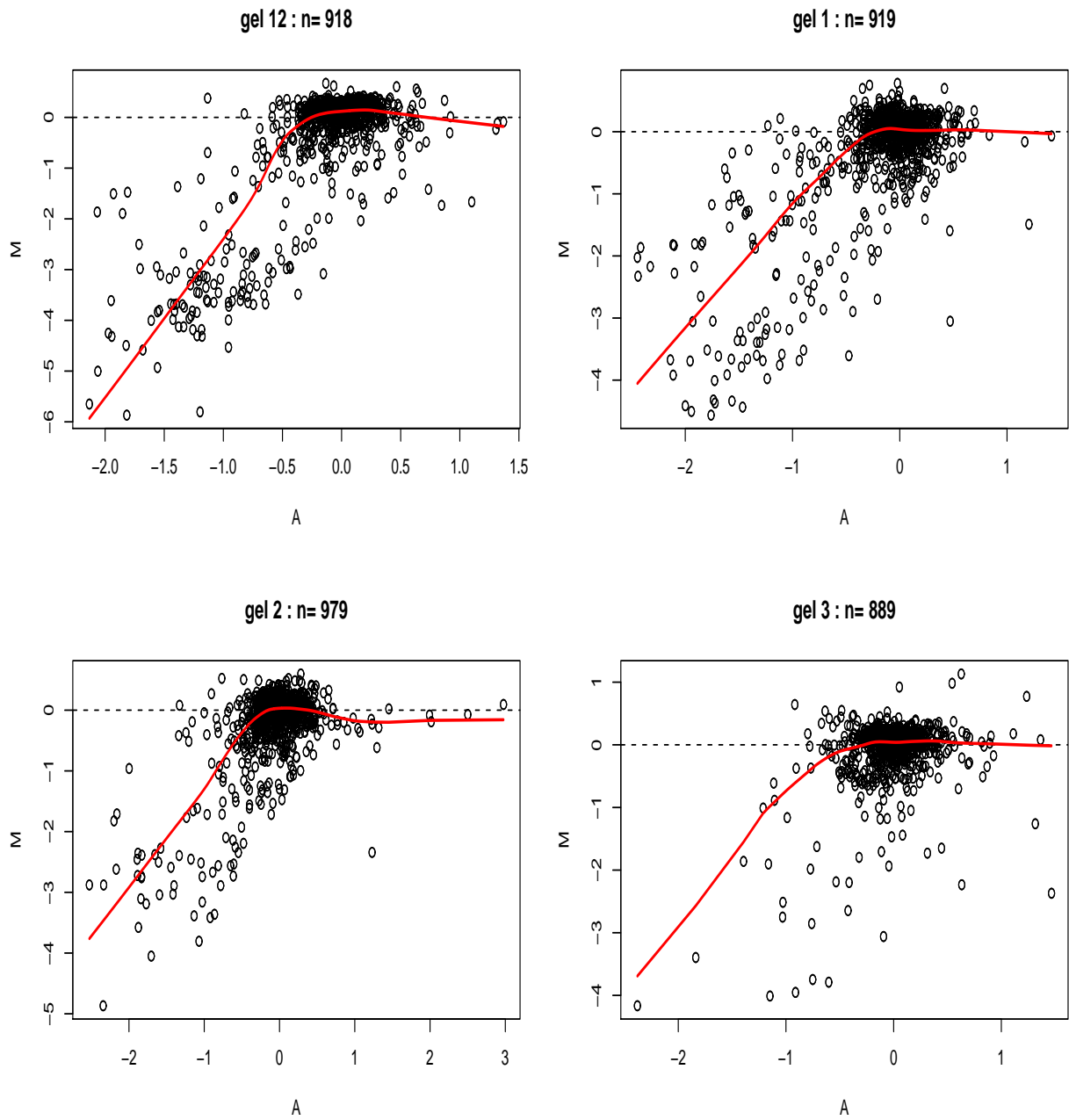


Figure 6:  $M - A$  plots for four gels in the PrecastNoFiltLandmrk experiment corresponding to the normalization method in Section 3.3.1.

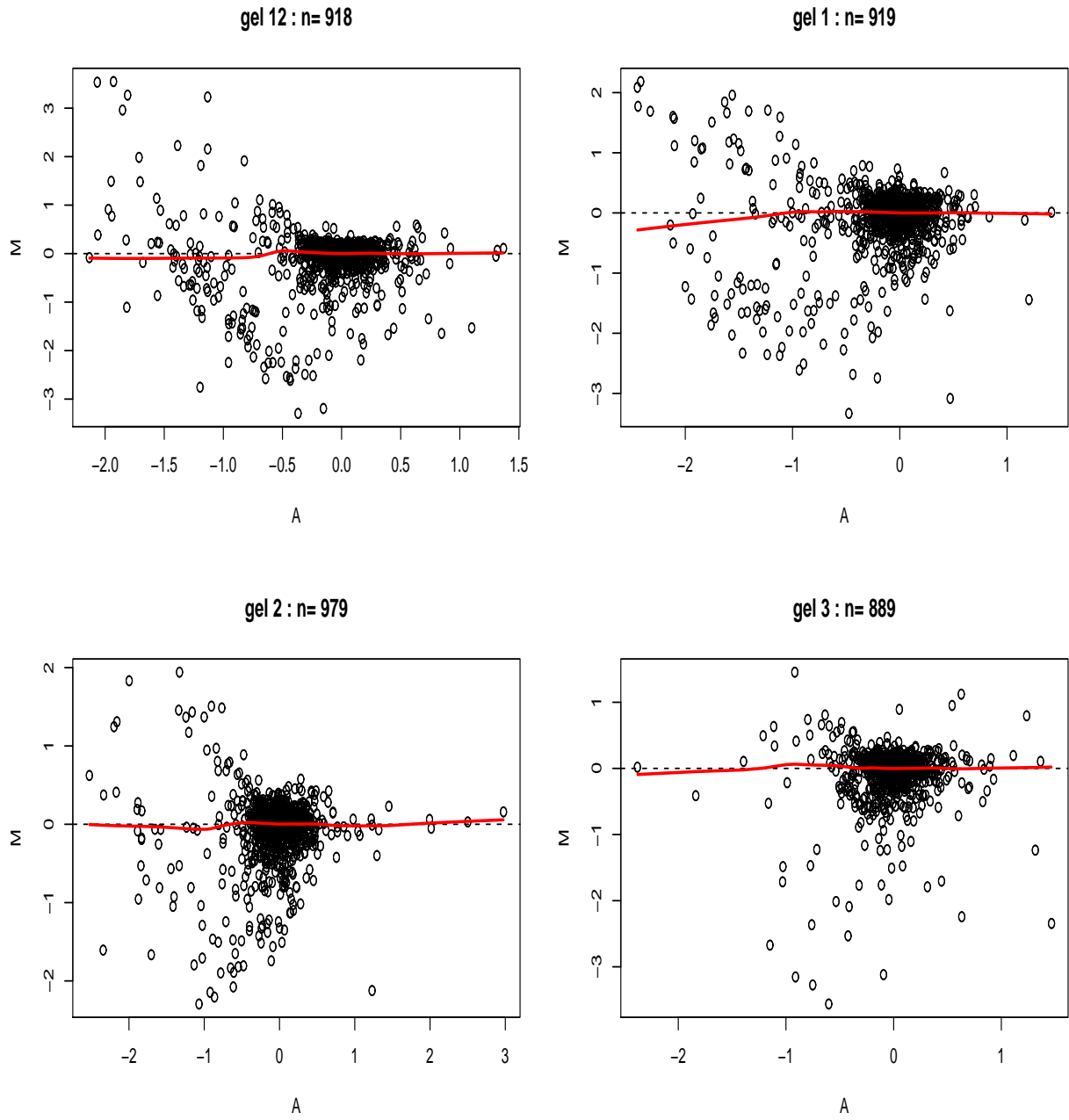
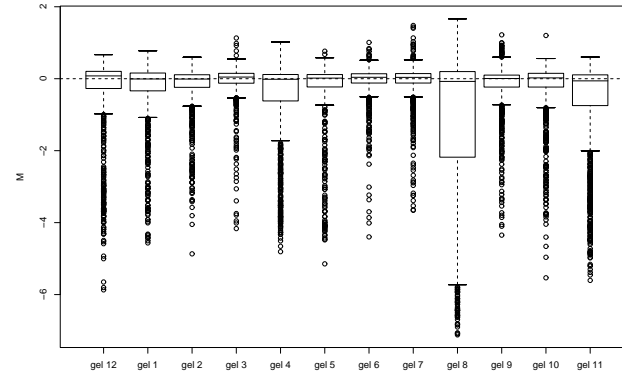
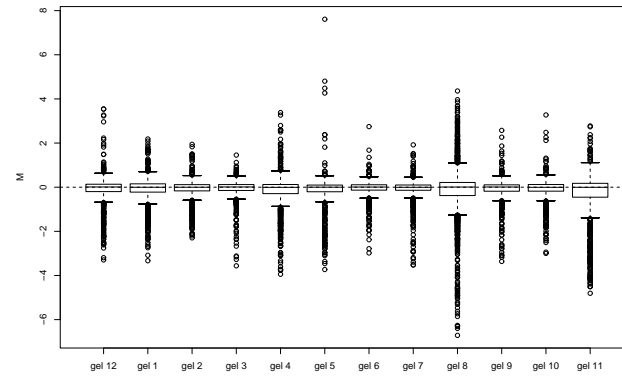


Figure 7: The normalized  $\tilde{M} - \tilde{A}$  plots for four gels in the PrecastNoFiltLandmrk experiment corresponding to the normalization method in Section 3.3.1.

(a)



(b)



(c)

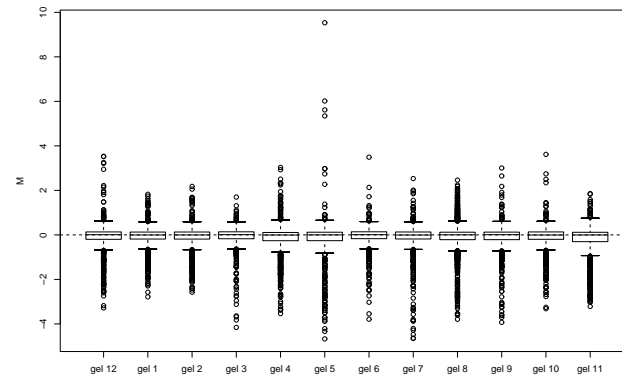


Figure 8: Boxplots of the M-values for the twelve gels from the PrecastNoFiltLandmrk dataset based on the: (a) original standardized log abundances, (b) within-gel normalized values using the method in Section 3.3.1, and (c) within- and between-gel normalized standardized log abundances using the method in Section 3.3.1.

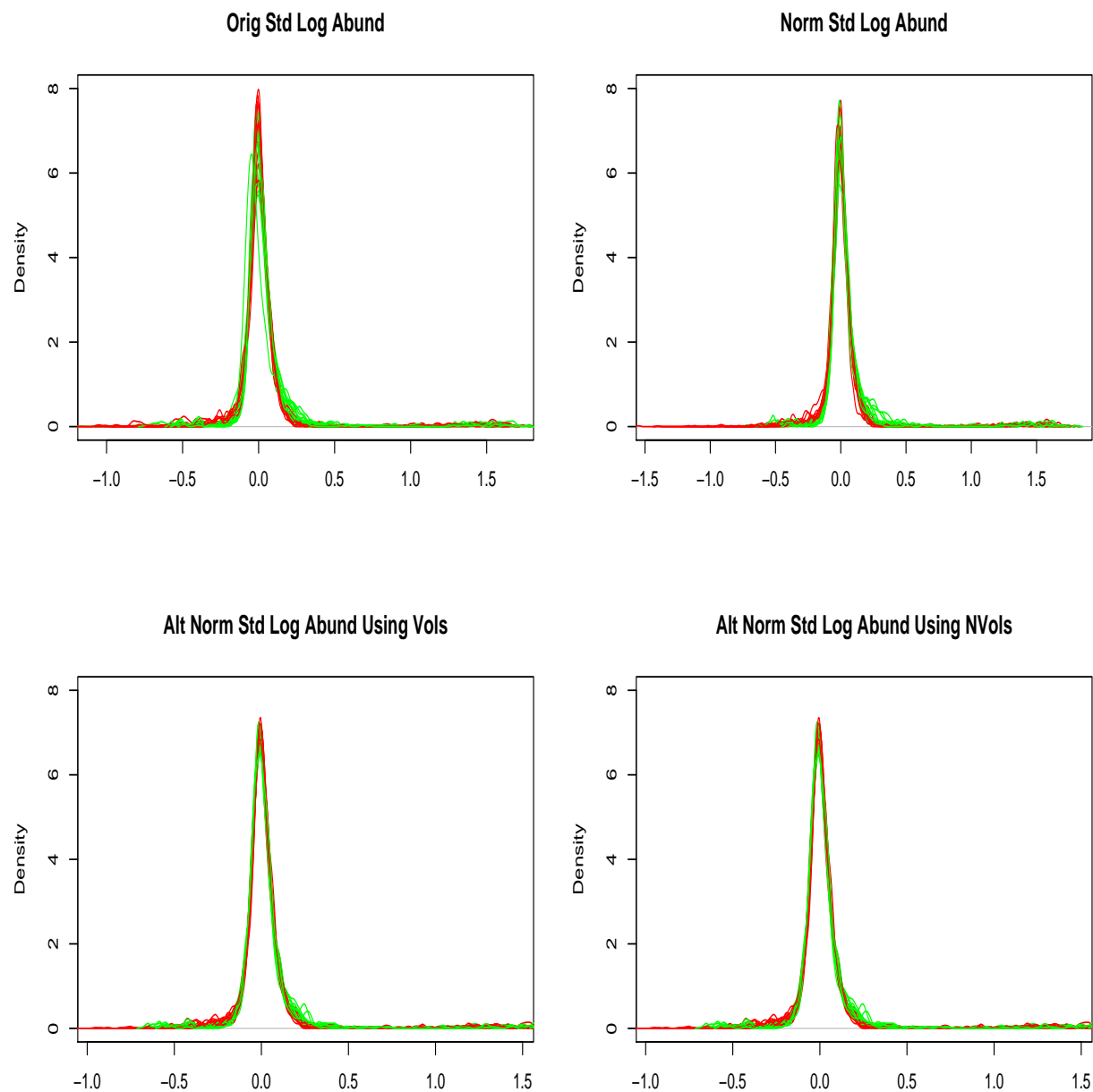


Figure 9: Density estimates of the standardized log abundances from the PrecastNoFiltLandmrk dataset: (a) original standardized log abundances from DeCyder, (b) normalized with the method in Section 3.3.1, (c) normalized with the method in Section 3.3.3, and (d) normalized with the method in Section 3.3.2. Two lines for each of the twelve gels corresponding to the distributions of *SLA32* (green) and *SLA52* (red).

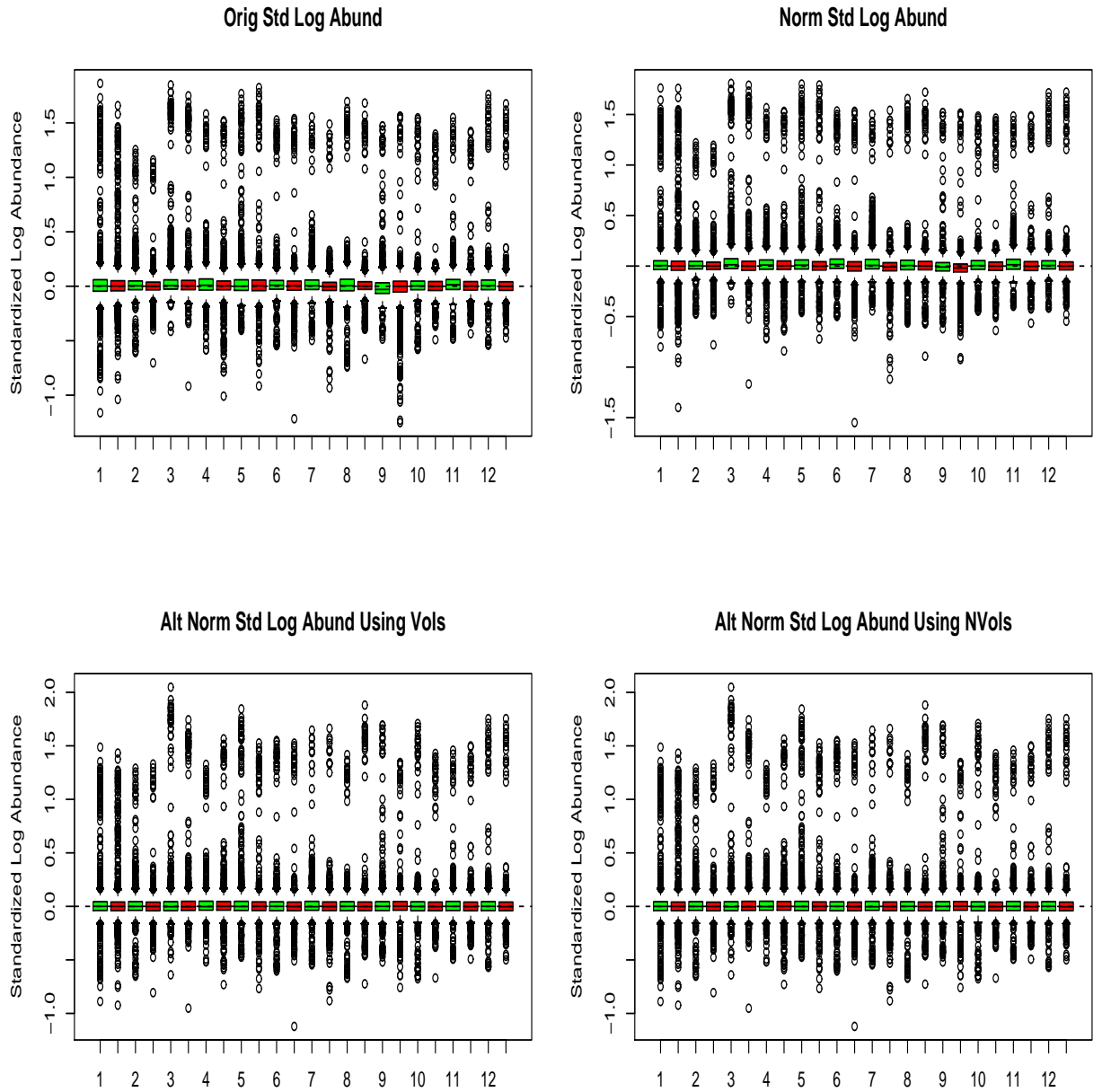


Figure 10: Boxplots of the standardized log abundances from the PrecastNoFiltLandmrk dataset: (a) original standardized log abundances from DeCyder, (b) normalized with the method in Section 3.3.1, (c) normalized with the method in Section 3.3.3, and (d) normalized with the method in Section 3.3.2. Two lines for each of the twelve gels corresponding to the distributions of *SLA32* (green) and *SLA52* (red).

Table 14: 95% prediction intervals for the expression ratios on a future gel. Columns indicate the subsets of the spots used in the calculations, rows the dataset used. *SLA*: based on the original standardized log abundance data from DeCyder; *SLA*<sup>(1)</sup>: based on the additionally normalized data using the standardized log abundances in Section 3.3.1; *SLA*<sup>(2)</sup>: based on the additionally normalized data using the normalized volumes in Section 3.3.2; *SLA*<sup>(3)</sup>: based on the additionally normalized data using the original volumes in Section 3.3.3.

Data	All spots	Spots matched on 8+ gels
<i>SLA</i>	[-2.55, 6.87]	[-1.84, 1.90]
<i>SLA</i> <sup>(1)</sup>	[-2.14, 7.50]	[-1.61, 1.99]
<i>SLA</i> <sup>(2)</sup>	[-1.99, 5.95]	[-1.53, 1.80]
<i>SLA</i> <sup>(3)</sup>	[-1.99, 5.95]	[-1.53, 1.80]

prediction interval can be constructed for the 97.5th percentiles. Since the distributions of the standardized log abundances corresponding to the two dyes on the same gel did not exhibit systematic differences, they were combined in the calculation of the prediction intervals. The effective sample size was then doubled to  $n = 24$ .

The first row and column of Table 14 presents the combined 95% prediction intervals for the 2.5th and 97.5th percentiles, based on the standardized log abundances for all the spots from the PrecastNoFiltLandmrk data. The interval provides an estimate on how large expression ratio values will likely have to be in order to correspond to real effects in future experiments similar to this one. According to the results, a biological sample will have to be down-regulated about 2.5-fold or up-regulated by more than 6-fold to be significant at the 5% level.

The second, third, and fourth rows present the corresponding results based on the standardized log abundances that were additionally normalized using the methods in Sections 3.3.1, 3.3.2, and 3.3.3, respectively. The second and third normalization methods led to identical results up to two decimal points, and reduced the length of the previous prediction interval.

The prediction limits in the first column of Table 14 can be tightened by considering certain subsets of the spots. The second column contains the corresponding re-

sults when including only the spots that were matched on at least eight of the twelve gels.

Many publications using DIGE report 1.2-fold change differences as biologically significant. A careful reading of the main DIGE reference paper on which the 1.2-fold change requirement is based on [13] reveals that the claim that DIGE can detect a 1.2-fold change is valid only for a subset of the spots (spot volumes in a certain range). Table 6 of [13] is similar to Table 14 here. For the best subset of spots, the lower and upper limits of the prediction intervals in Table 6 of [13] are  $[-1.2, 1.2]$ . In addition to the best-case scenario, Table 6 of [13] presents additional intervals for different subsets of spots. Note that a fold-change as large as 29.1 is also part of that table, even for the simpler biological samples (mouse liver homogenates) studied in that paper.

Table 15 summarizes the distributions of the spot-wise standard deviations of the three additionally normalized datasets. All three normalization methods reduced the spot-wise standard deviations across the gels, as evidenced by comparing Table 13 to Table 15. In general, the method in Section 3.3.1 resulted in the largest improvement, but the results based on the methods in Sections 3.3.2 and 3.3.3 were close. For the combined *SLA*32 and *SLA*52 data using the spots matched on at least 10 gels, the standard deviation reduction from 0.0422 in Table 13 to 0.0391 in Table 15 corresponds to a 14.15% reduction in the variance. Using only the spots matched on at least 70% of the gels, i.e. matched on at least 8 gels, the corresponding standard deviation reduction from 0.0427 to 0.0406 is equivalent to a 9.60 percent reduction in the variance.

The three additional normalization methods led to more dramatic decreases in the standard deviations in the other four datasets described in Section 3. Our interpretation is that the better quality the gels are, the less profound effect the extra normalization steps have. However, if resources prohibit repeating experiments that have relatively poor quality gels, the extra normalization steps can mitigate the problems and increase the confidence in the results.

To investigate the effect of normalization on the identification of “differentially expressed” spots, a one-fay fixed effect Analysis of Variance (ANOVA) model was fitted first to the original standardized log abundances, then to the additionally normalized standardized log abundances. Assuming no dye or gel effects, the goal was to



Table 15: Normalized PrecastNoFiltLandmrk dataset: summary statistics for the standard deviations of the normalized standardized log abundances based on three different normalizations. Caption as in Table 5. For each subset, the first row is based on normalizing the standardized log abundances (Section 3.3.1), the second on normalizing the the normalized volumes (Section 3.3.2), the third on normalizing the original volumes (Section 3.3.3).

Data	MatchStatus	NoSpots	Min	1stQu	Med	Mean	3rdQu	Max
SLA32	10+	549	0.0081	0.0277	0.0372	0.0434	0.0507	0.2019
			0.0097	0.0282	0.0368	0.0435	0.0516	0.2831
			0.0097	0.0283	0.0368	0.0435	0.0516	0.2832
	8+	1055	0.0081	0.0281	0.0383	0.0471	0.0548	0.4135
			0.0081	0.0284	0.0390	0.0473	0.0549	0.4312
			0.0081	0.0284	0.0390	0.0473	0.0549	0.4313
	6+	1575	0.0069	0.0291	0.0405	0.0525	0.0596	0.5598
			0.0081	0.0295	0.0414	0.0531	0.0593	0.5907
			0.0081	0.0295	0.0414	0.0531	0.0593	0.5908
	4+	2041	0.0054	0.0299	0.0428	0.0590	0.0645	0.6137
			0.0072	0.0302	0.0436	0.0584	0.0648	0.6381
			0.0072	0.0302	0.0435	0.0584	0.0648	0.6377
	2+	2414	0.0005	0.0293	0.0434	0.0643	0.0697	0.8786
			0.0002	0.0300	0.0440	0.0629	0.0666	0.7846
			0.0002	0.0300	0.0440	0.0629	0.0666	0.7836
SLA52	10+	549	0.0142	0.0300	0.0384	0.0443	0.0520	0.1853
			0.0087	0.0298	0.0372	0.0447	0.0528	0.2130
			0.0087	0.0298	0.0372	0.04477	0.0528	0.2130
	8+	1055	0.0092	0.0291	0.0388	0.0464	0.0545	0.4448
			0.0080	0.0288	0.0381	0.0459	0.0537	0.4178
			0.0080	0.0288	0.0380	0.0458	0.0537	0.4180
	6+	1575	0.0081	0.0295	0.0403	0.0507	0.0579	0.6985
			0.0080	0.0290	0.0390	0.0497	0.0560	0.6548
			0.0080	0.0290	0.0390	0.0496	0.0560	0.6547
	4+	2041	0.0060	0.0300	0.0418	0.0565	0.0626	0.8264
			0.0029	0.0292	0.0402	0.0540	0.0592	0.7643
			0.0028	0.0292	0.0402	0.0540	0.0593	0.7641
	2+	2414	0.0011	0.0295	0.0425	0.0628	0.0660	1.0260
			0.0000	0.0286	0.0405	0.0588	0.0628	1.0910
			0.0000	0.0286	0.0405	0.0588	0.0628	1.0910
SLA32 and SLA52	10+	549	0.0164	0.0312	0.0391	0.0474	0.0527	0.1944
			0.0136	0.0314	0.0391	0.0475	0.0538	0.2222
			0.0135	0.0314	0.0391	0.0475	0.0538	0.2223
	8+	1055	0.0160	0.0310	0.0406	0.0517	0.0580	0.4166
			0.0136	0.0313	0.0411	0.0516	0.0576	0.4132
			0.0135	0.0313	0.0411	0.0516	0.0576	0.4133
	6+	1575	0.0115	0.0318	0.0426	0.0579	0.0637	0.6042
			0.0136	0.0323	0.0431	0.0574	0.0630	0.5961
			0.0135	0.0323	0.0431	0.0574	0.0630	0.5960
	4+	2041	0.0115	0.0327	0.0453	0.0661	0.0733	0.6797
			0.0135	0.0333	0.0460	0.0642	0.0700	0.6549
			0.0135	0.0333	0.0460	0.0642	0.0700	0.6547
	2+	2414	0.0049	0.0331	0.0468	0.0733	0.0784	0.7692
			0.0124	0.0338	0.0474	0.0703	0.0758	0.7959
			0.0124	0.0338	0.0474	0.0703	0.0758	0.7953

Table 16: The number of spots with greater than 1.5 fold-change between any two samples, and with ANOVA sample effect p-value less than 0.05, as a function of the subset of the spots and the type of p-value used, in the format is  $n_1/n_2$ , where  $n_1$  corresponds to the standardized log abundance data, and  $n_2$  to the additionally normalized standardized log abundance data. Total number of spots: 2511. Number of spots matched on 8+ gels: 1055.

	All spots	Matched on 8+
p-value	111/74	53/29
FDR-adj p-value	33/16	36/17

investigate the number spots that showed differential expression among the 8 samples A through H. Per the experimental design in Figure 1, three replicates were available from each of the 8 samples. Since the samples were biologically identical on each of the gels, the observed differentially expressed spots provide a measure of the technical variability.

Table 16 summarizes the results in terms of the number of spots that satisfied the following criterion: at least 1.5 fold-change between any two samples of the eight and ANOVA p-value for sample effect less than 0.05. The two rows indicate whether unadjusted or FDR-adjusted p-values were used. The columns differentiate the type of data used when fitting the model (original standardized log abundances from DeCyder or the additionally normalized standardized log abundances using the normalization method in Section 3.3.1), and the subset of the data used (whether all spots were included or only the spots matched on at least 8 gels). The additional normalization decreased the number of potential spots to consider in every case. In the most relevant case of using only spots matched on at least 8 gels and the adjusted p-value, the reduction in false positives is over 50% (17 vs 36).

## 5 Summary

We described and evaluated several statistical methods to normalize the data across different dyes and gels in complex 2-D DIGE experiments. Algorithms from the mi-

croarray analysis community provided the basis of the methods. We demonstrated the effects of the proposed methods in correcting for biases observed in the protein expressions measured with different dyes on the same gel, as well as in correcting inconsistencies observed in protein expressions due to differences in the gels. As a result of the additional normalizations, the spot-wise variances across the gels were reduced by about 10%, which increased the reliability of the system and the confidence in the methods. The new protein expression normalization methods coupled with the statistically necessary p-value adjustment reduced the number of potentially differentially expressed spots by more than half.

A balanced design instead of Table 1 would have been more efficient in testing the effects of filtering and land-marking but was not possible due to limited resources. Although the present study did not find compelling evidence for the use of filtering, a better designed study using the best quality gels and a carefully selected combination of filter thresholds may determine the optimal set of filters.

Continued advances in 2-D DIGE gel technology are improving the quality of the data, and the viability of the platform to detect differential protein expression in biological samples with confidence, and in a reproducible manner. The quality of precast gels has improved considerably since the first experiment reported in this study. At that time, precast gels were not reliable, and thus experimenters had to pour their own gels. Based on our findings, the newly available precast gels are more consistent than gels poured in the lab, and are thus preferable for use in future experiments.

In addition to technological advances, the quality of the DeCyder software is also maturing. The latest release (version 6.0) includes an improved spot detection and matching algorithm, as well as p-value correction for multiple hypothesis testing.

Areas where further improvements over the existing methods would benefit the 2-D DIGE community include experimental design, image processing, and statistical modeling and testing to assess the differential expression of proteins in complicated experiments.

The methods evaluated here are implemented in software and are readily available for use in 2-D DIGE proteomics experiments throughout the lab. Use of the advocated statistical procedures adds more rigor to the analysis and interpretation 2-D DIGE experiments.

## Acknowledgments

We thank the Proteomics group of Dr. Sandra L. McCutchen-Maloney, and especially Todd H. Corzett, for providing the gels, the description of the Sample preparation and gel processing section, and assistance with the analyses. Funding from the Computation Directorate Technology Base Portfolio is gratefully acknowledged. This work was performed under the auspices of the U.S. Department of Energy by University of California Lawrence Livermore National Laboratory under Contract No. W-7405-Eng-48, UCRL.

## References

- [1] A. Alban, S. O. David, L. Bjorkesten, C. Andersson, E. Sloge, S. Lewis, and I. Currie. A novel experimental design for comparative two-dimensional gel analysis: Two-dimensional difference gel electrophoresis incorporating a pooled internal standard. *Proteomics*, 3:36–44, 2003.
- [2] Amersham Biosciences. *DeCyder Differential Analysis Software User Manual, Version 5.0*, 2003.
- [3] Y. Benjamini and Y. Hochberg. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society, Series B*, 57:289–300, 1995.
- [4] X. Cui and G. A. Churchill. Statistical tests for differential expression in cDNA experiments. *Genome Biology*, 4:210.a–210.9, 2003.
- [5] S. Dudoit and Y. H. Yang. *The Analysis of Gene Expression Data: Methods and Software*, chapter Bioconductor R packages for exploratory analysis and normalization of cDNA microarray data, pages 73–101. Springer, New York, 2003.
- [6] A. Görg, C. Obermaier, G. Boguth, A. Harder, B. Scheibe, R. Wildgruber, and W. Weiss. The current state of two-dimensional electrophoresis with immobilized pH gradients. *Electrophoresis*, 21(6):1037–53, Apr 2000.
- [7] W. Huber, A. von Heydebreck, H. Sülthmann, A. Poustka, and M. Vingron. Variance stabilization applied to microarray data calibration and to the quantification of differential expression. *Bioinformatics*, 18:S96–S104, 2002.
- [8] W. Huber, A. von Heydebreck, and M. Vingron. *Handbook of Statistical Genetics*, chapter Analysis of Microarray Gene Expression Data. Wiley, 2nd edition, 2003.
- [9] K. S. Lilley, A. Razzaq, and P. Dupree. Two-dimensional gel electrophoresis: recent advances in sample preparation, detection and quantitation. *Curr. Opin. Chem. Biol.*, 6(1):46–50, Feb 2002.
- [10] P. H. O’Farrell. High resolution two-dimensional electrophoresis of proteins. *J. Biol. Chem.*, 250:4007–4021, 1975.
- [11] J. P. Shaffer. Multiple hypothesis testing. *Ann. Rev. Psych.*, 46:561–576, 1995.
- [12] G. K. Smyth, Y. H. Yang, and T.P. Speed. *Methods in Molecular Biology*, volume 224, chapter Statistical Issues in cDNA Microarray Data Analysis, pages 111–136. Humana Press, Totowa, NJ, 2003.
- [13] R. Tonge, J. Shaw, B. Middleton, R. Rowlinson, S. Rayner, J. Young, F. Pognan, E. Hawkins, I. Currie, and M. Davidson. Validation and development of fluorescence two-dimensional gel electrophoresis proteomics technology. *Proteomics*, 1:377–396, 2001.
- [14] M. Ünlü, M. E. Morgan, and J. S. Minden. Difference gel electrophoresis: A single gel method for detecting changes in protein extracts. *Electrophoresis*, 18:2071–2077, 1997.